Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2035

Air Quality Index Prediction using Bi-LSTM

Y V Nagesh Meesala[1]
Associate Professor
Department of CSE, Raghu Engineering College,
Visakhapatnam, Andhra Pradesh, India

Avanapu Uma Mahesh[2]; Abhiram Reddy Bellana[3]
Bichukathula Obulesu[4]
Students
Department of CSE, Raghu Engineering College,
Visakhapatnam, Andhra Pradesh, India

Abstract:- Air quality prediction plays a vital role in safeguarding public health and guiding environmental policy. Traditional single-model approaches often struggle to accurately forecast air quality fluctuations. In response, this study introduces a robust prediction system leveraging advanced machine learning techniques. We present a comparative analysis of several models including Support Vector Regression (SVR), Genetic Algorithm-Enhanced Extreme Learning Machine (GA-KELM), and Deep Belief Network with Back-Propagation (DBN-BP). Additionally, we propose the integration of Bidirectional Long Short-Term Memory (BiLSTM), a deep learning architecture, to further enhance prediction accuracy. Through comprehensive experimentation and evaluation, we demonstrate that BiLSTM outperforms existing models, exhibiting lower Root Mean Square Error (RMSE) and Mean Squared Error (MSE) values. Furthermore, by incorporating GA-KELM, we optimize the performance of BiLSTM, enhancing its predictive capabilities even further. The proposed hybrid model not only offers improved accuracy in air quality forecasting but also contributes to informed decision-making for pollution control strategies and public health interventions. This research underscores the significance of exploring innovative techniques to address pressing environmental challenges and underscores the potential of machine learning in advancing air quality management.

Keywords:- *Time Series, Air Quality Forecasting, Machine Learning, Extreme Learning Machine, Genetic Algorithm*.

I. INTRODUCTION

Air pollution has emerged as a pressing global concern in the twenty-first century, exacerbated by rapid industrialization and urbanization [1]. The consequences of deteriorating air quality are far-reaching, impacting both the environment and public health [2]. Research by Li et al. underscores the health risks associated with outdoor physical activity in the presence of ambient air pollution, particularly in regions experiencing rapid industrial growth such as China [3]. In China, as in many other countries, air quality is measured using parameters outlined in the Chinese Ambient Air Quality Standards, including sulfur dioxide (SO2), nitrogen dioxide (NO2), particulate matter with sizes less than 10 microns (PM10) and 2.5 microns (PM2.5), ozone (O3), and carbon monoxide (CO) [4].

The adverse effects of these pollutants on human health are well-documented [5]. The International Energy Agency estimates that air pollution contributes to approximately 6.5 million premature deaths annually, with long-term exposure to pollutants like PM2.5 and traffic-related emissions linked to higher incidences of lung cancer, coronary heart disease, and other illnesses [6]. Consequently, there is a growing urgency to develop effective strategies for air quality prediction, which is integral to environmental protection efforts [7].

Air quality prediction relies heavily on data collected from monitoring stations scattered across major cities [8]. These stations provide valuable insights into pollution levels and help inform predictive models. Machine learning algorithms have emerged as powerful tools for analyzing such data, offering the ability to automatically learn features at multiple levels of abstraction [9]. However, challenges persist, including the limited availability of comprehensive datasets and the complexity of modeling multiple pollutants simultaneously [10].

Recent research has explored various approaches to address these challenges. Wu Q. et al. proposed an optimal-hybrid model for daily Air Quality Index (AQI) prediction, leveraging data from six atmospheric pollutants [11]. However, traditional neural network algorithms often encounter issues such as slow learning, susceptibility to local minima, and complex training processes [12].

To overcome these limitations, Huang et al. introduced the extreme learning machine (ELM) algorithm, which is based on the generalized inverse matrix theory and features a single hidden layer feedforward neural network [13]. The ELM algorithm has demonstrated superior performance in AQI prediction compared to traditional neural networks, offering advantages in terms of parameter selection, training speed, and prediction accuracy [14]. Despite its effectiveness, the ELM algorithm's reliance on randomly selected parameters for hidden layer nodes poses challenges to prediction accuracy [15].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2035

In this context, this paper aims to address the shortcomings of existing air quality prediction models by proposing a novel approach that combines the strengths of machine learning algorithms with enhanced parameter optimization techniques. Specifically, we introduce a hybrid model that integrates the Bidirectional Long Short-Term Memory (BiLSTM) architecture with the Genetic Algorithm-Enhanced Extreme Learning Machine (GA-KELM). This combination aims to improve the accuracy and robustness of air quality predictions by leveraging the predictive capabilities of BiLSTM while optimizing model parameters through genetic algorithms [16].

In summary, this paper contributes to the ongoing efforts to advance air quality prediction methodologies by introducing a novel hybrid model that addresses the limitations of existing approaches. By combining BiLSTM and GA-KELM, we aim to provide more accurate and reliable predictions, thereby facilitating informed decision-making for environmental protection and public health management.

II. LITERATURE SURVEY

Air pollution has emerged as a significant environmental and public health issue globally, necessitating comprehensive research to understand its causes, effects, and mitigation strategies. In this literature survey, we review key studies related to air pollution monitoring, forecasting, and control, with a focus on the application of machine learning techniques for air quality prediction.

Li et al. (2019) highlighted air pollution as a global problem that requires local solutions, emphasizing the importance of addressing air quality issues at the regional level [1]. This perspective underscores the need for localized air quality monitoring and prediction systems to inform targeted interventions. Han et al. (2018) introduced a Bayesian Long Short-Term Memory (LSTM) model to evaluate the effects of air pollution control regulations in China, demonstrating the utility of advanced statistical techniques for analyzing air quality data [2]. Their work highlights the potential of LSTM models in predicting the impacts of policy interventions on air quality outcomes.

Bai et al. (2018) provided an overview of air pollution forecasts, discussing various modeling approaches and data sources used in air quality prediction [3]. Their review underscores the complexity of air quality forecasting and the importance of incorporating diverse data sources, including meteorological data, satellite observations, and ground-level monitoring data. Ding and Xue (2019) proposed a deep learning approach for writer identification using inertial sensor data of air handwriting, demonstrating the versatility of deep learning techniques in analyzing sensor data for diverse applications [4].

Cheng et al. (2019) investigated the optimization of outdoor air ratio in air conditioning systems for achieving targeted indoor air quality and maximal energy savings [5]. Their study highlights the importance of optimizing ventilation strategies to maintain indoor air quality while minimizing energy consumption. Chaudhary et al. (2018) developed a time series-based LSTM model to predict air pollutant concentrations in prominent cities in India, showcasing the applicability of LSTM models in air quality forecasting [6]. Their research contributes to the growing body of literature on data-driven approaches to air quality prediction.

Chen et al. (2018) proposed an urban healthcare big data system based on crowdsourced and cloud-based air quality indicators, illustrating the potential of crowdsourcing data for monitoring urban air quality [7]. Their study highlights the role of emerging technologies in expanding the scope of air quality monitoring and public health surveillance. Du et al. (2021) presented a deep air quality forecasting framework using a hybrid deep learning approach, combining convolutional neural networks (CNNs) and LSTM networks [8]. Their research demonstrates the effectiveness of hybrid deep learning models in capturing complex spatiotemporal patterns in air quality data.

Overall, the literature survey highlights the growing interest in leveraging machine learning techniques for air quality monitoring, forecasting, and management. Studies have explored a wide range of approaches, including LSTM models, deep learning frameworks, and hybrid machine learning architectures, to improve the accuracy and reliability of air quality predictions. These advancements hold promise for informing evidence-based interventions to mitigate the adverse effects of air pollution on public health and the environment.

III. METHODOLOGY

A. Proposed Work:

The proposed work aims to integrate Genetic Algorithm (GA) with Extreme Learning Machine (ELM) to enhance air quality prediction, with a specific focus on predicting PM2.5 levels. GA will be employed to optimize the selection of hidden nodes and layers within the ELM model, thereby improving its learning capability and prediction accuracy. By leveraging GA's[14] ability to search for optimal solutions within a predefined search space, the ELM model can adaptively adjust its architecture to better capture the complex relationships inherent in air quality data. Comparative analysis will be conducted against traditional methods such as Support Vector Machines (SVM)[16], with performance metrics including Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) used to evaluate effectiveness. The proposed approach seeks to provide a more robust and accurate air quality prediction system, facilitating comprehensive assessments of pollution levels and their potential impact on public health and the environment.

B. System Architecture:

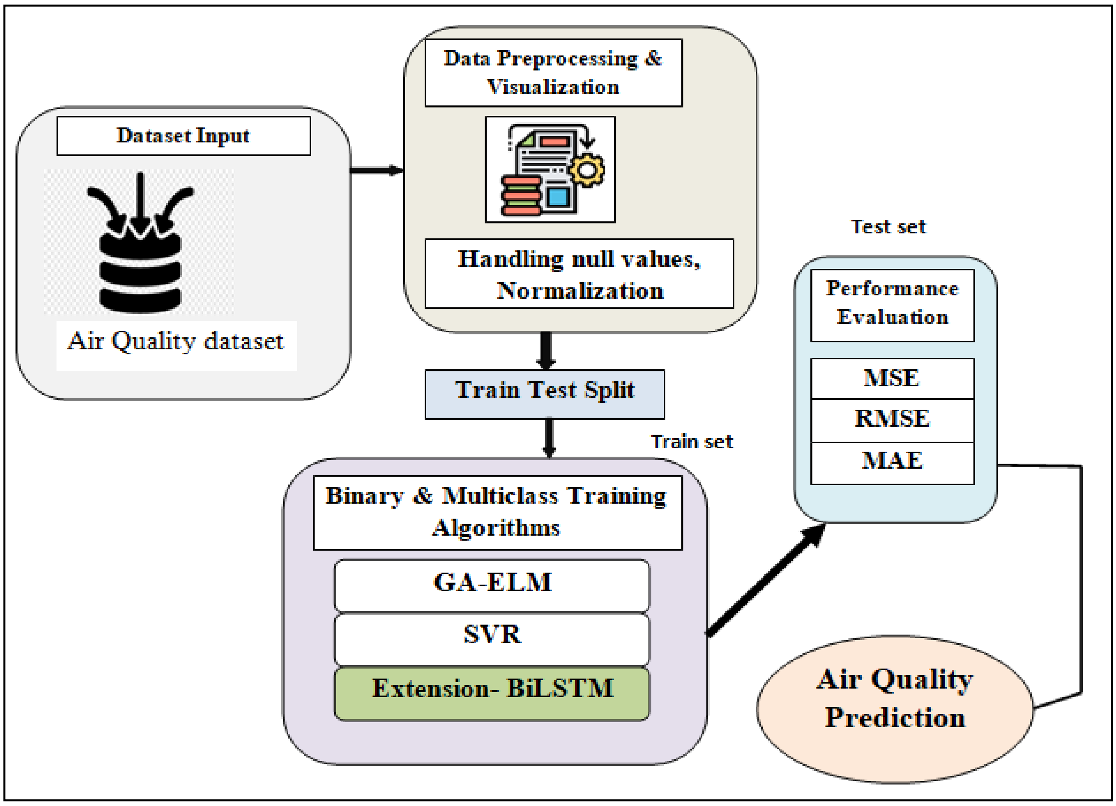

Fig 1 Proposed Architecture

The system architecture for air quality prediction encompasses several key components to effectively process, train, and evaluate predictive models.

➢ *Dataset Input:*

The system begins by ingesting air quality datasets containing relevant features such as pollutant concentrations, meteorological data, and geographical information.

➢ *Data Processing and Visualization:*

Pre-processing steps include handling null values, normalization, and feature engineering to prepare the data for modeling. Visualization techniques are employed to gain insights into data distributions and correlations.

➢ *Train-Test Split:*

The dataset is divided into training and testing sets to facilitate model training and evaluation. This ensures that the model's performance is assessed on unseen data, helping to gauge its generalization ability.

➢ *Binary and Multi-Class Training Algorithms:*

The system incorporates various training algorithms, including Genetic Algorithm-Enhanced Extreme Learning Machine (GA-ELM)[14], Support Vector Regression (SVR)[16], and Bidirectional Long Short-Term Memory (BiLSTM) networks. These algorithms are trained on the training data to learn the underlying patterns and relationships between input features and air quality outcomes.

➢ *Performance Evaluation:*

Model performance is evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics quantify the discrepancies between predicted and actual air quality values, providing insights into the accuracy and reliability of the models.

➢ *Air Quality Prediction:*

Once trained, the models are deployed to make predictions on unseen data, estimating air quality parameters such as pollutant concentrations or Air Quality Index (AQI) values. These predictions are crucial for assessing current and future air quality conditions, enabling informed decision-making for pollution control and public health interventions.

Overall, the system architecture provides a comprehensive framework for air quality prediction, leveraging machine learning algorithms and performance evaluation techniques to deliver accurate and reliable predictions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2035

C. Dataset:

Table 1 Dataset

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
0	Ahmedabad	2015-01-01	0.0	0.0	0.92	18.22	17.15	0.0	0.92	27.64	133.36	0.00	0.02	0.00	0.0	0
1	Ahmedabad	2015-01-02	0.0	0.0	0.97	15.69	16.46	0.0	0.97	24.55	34.06	3.68	5.50	3.77	0.0	0
2	Ahmedabad	2015-01-03	0.0	0.0	17.40	19.30	29.70	0.0	17.40	29.07	30.70	6.80	16.40	2.25	0.0	0
3	Ahmedabad	2015-01-04	0.0	0.0	1.70	18.48	17.97	0.0	1.70	18.59	36.08	4.43	10.14	1.00	0.0	0
4	Ahmedabad	2015-01-05	0.0	0.0	22.10	21.42	37.76	0.0	22.10	39.33	39.31	7.01	18.89	2.78	0.0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
497	Ahmedabad	2016-05-12	0.0	0.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0
498	Ahmedabad	2016-05-13	0.0	0.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0
499	Ahmedabad	2016-05-14	0.0	0.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0
500	Ahmedabad	2016-05-15	0.0	0.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0
501	Ahmedabad	2016-05-16	0.0	0.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0

The air quality dataset comprises measurements of various pollutants such as sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter with sizes less than 10 microns (PM10) and 2.5 microns (PM2.5), ozone (O_3), and carbon monoxide (CO). Each observation includes pollutant concentrations, along with corresponding timestamps and geographical locations. Additionally, meteorological data such as temperature, humidity, wind speed, and atmospheric pressure may be included. This dataset enables exploration and analysis of air quality trends over time and across different regions, facilitating research on the impact of pollution on public health and the environment.

D. Data Processing:

➢ *Data Processing with Pandas DataFrame:*

The Pandas DataFrame is utilized for efficient data manipulation and preprocessing tasks. This includes handling missing values, normalization, and dropping unwanted columns to prepare the dataset for model training.

- *Handling Missing Values:*

Missing values, if any, are addressed through techniques such as imputation or removal. This ensures the integrity of the dataset and prevents biases in subsequent analyses.

- *Normalization:*

Numeric features are normalized to a standard scale, typically between 0 and 1, to ensure consistency and prevent features with larger scales from dominating the model training process.

- *Dropping Unwanted Columns:*

Columns that are irrelevant or redundant for the predictive task are dropped from the DataFrame. This reduces dimensionality and enhances computational efficiency during model training.

➢ *Data Processing with Keras Data Frame:*

The Keras DataFrame facilitates seamless integration with deep learning frameworks, enabling efficient data preprocessing and model training for neural network architectures.

- *Handling Missing Values:*

Similar to Pandas, missing values are addressed to ensure data completeness and integrity.

- *Normalization:*

Numeric features are normalized within the Keras DataFrame using appropriate scaling techniques. This ensures consistent feature ranges and aids in convergence during neural network training.

➢ *Dropping Unwanted Columns:*

Columns deemed unnecessary for neural network training are dropped from the Keras DataFrame. This optimizes computational resources and prevents overfitting by reducing model complexity.

Overall, data processing with both Pandas and Keras DataFrames plays a crucial role in preparing the dataset for model training, ensuring data quality, and facilitating efficient model convergence.

E. Visualization:

Visualization using Seaborn and Matplotlib enhances understanding of air quality data through insightful graphical representations.

➢ *Histograms:*

Seaborn's `histplot` and Matplotlib's `hist` functions visualize the distribution of pollutant concentrations, revealing patterns and outliers.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2035

➢ *Scatter Plots:*

Seaborn's `scatterplot` and Matplotlib's `scatter` functions depict relationships between different pollutants or between pollutants and meteorological variables, aiding in identifying correlations.

➢ *Line Plots:*

Seaborn's `lineplot` and Matplotlib's `plot` functions display temporal trends in pollutant concentrations over time, facilitating the identification of seasonal variations and long-term trends.

These visualizations provide valuable insights into air quality dynamics, informing subsequent analysis and model development.

F. Feature Selection:

Feature selection is crucial for building effective air quality prediction models. Techniques such as correlation analysis, feature importance ranking, and dimensionality reduction methods like Principal Component Analysis (PCA) are employed. Correlation analysis identifies relationships between pollutants and meteorological variables, aiding in selecting relevant features. Feature importance ranking methods, such as Random Forest feature importances, prioritize influential features for prediction. Additionally, PCA identifies latent variables capturing the majority of data variance, reducing dimensionality while preserving essential information. By selecting the most informative features, feature selection optimizes model performance and computational efficiency in air quality prediction tasks.

G. Training & Testing:

Splitting the air quality dataset into training and testing subsets is essential for evaluating model performance. Typically, a random split, such as an 80/20 or 70/30 ratio, is applied, ensuring an adequate amount of data for both training and testing. The training set is used to train the predictive models, while the testing set remains unseen during training and is reserved for evaluating model performance. This split helps assess the model's ability to generalize to new data and ensures unbiased performance evaluation, thus enhancing the reliability of air quality predictions in real-world scenarios.

H. Algorithms:

➢ *Genetic Algorithm with Extreme Learning Machine (GA-ELM):*

The Genetic Algorithm with Extreme Learning Machine (GA-ELM) merges the evolutionary optimization capabilities of Genetic Algorithms (GAs)[14] with the efficient learning framework of Extreme Learning Machines (ELMs). In GA-ELM, the GA optimizes the parameters of the ELM model to enhance its predictive performance. The GA evolves a population of potential ELM solutions by iteratively selecting, crossing over, and mutating individuals based on their fitness, typically evaluated using a validation dataset. Meanwhile, the ELM employs a single hidden layer with random activation functions to map input features to a higher-dimensional space, followed by output weight calculation using the Moore-Penrose pseudoinverse.

➢ *Support Vector Regressor (SVR):*

Support Vector Regressor (SVR) employs the principle of structural risk minimization to fit a regression model. It aims to find the hyperplane that best separates data points while maximizing the margin. SVR[16] optimizes hyperparameters such as kernel type and regularization parameter during model training to minimize the loss function, typically epsilon-insensitive loss. Once trained, SVR utilizes the learned hyperplane to predict air quality values on unseen data, leveraging its ability to capture complex relationships between input features and output variables.

➢ *Bidirectional Long Short-Term Memory (BiLSTM):*

The extension of Bidirectional Long Short-Term Memory (BiLSTM) introduces a neural network architecture capable of capturing long-range dependencies in sequential data. BiLSTM processes input sequences in both forward and backward directions, allowing it to capture past and future context simultaneously. This capability is particularly useful for modeling temporal patterns in air quality data, where past and future observations may influence current air quality levels. BiLSTM models have demonstrated efficacy in capturing complex temporal dynamics, making them suitable for air quality prediction tasks.

IV. EXPERIMENTAL RESULTS

➢ *MSE:*

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD).

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

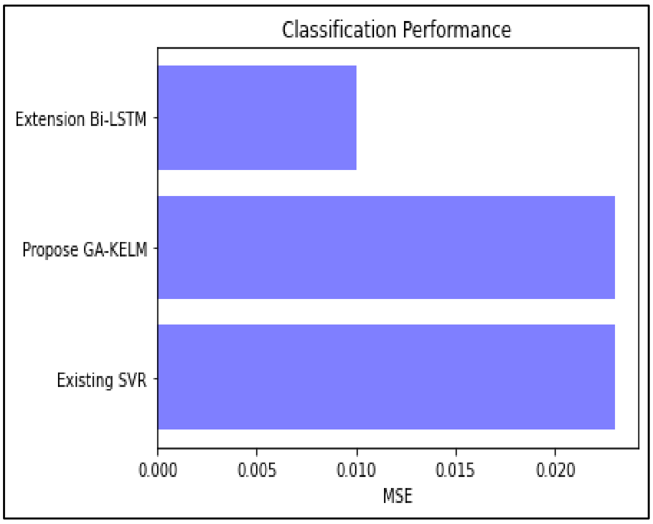

Fig 2 MSE Comparison Graph

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2035

➢ *RMSE:*

The root mean square error (RMSE) measures the average difference between a statistical model's predicted values and the actual values. Mathematically, it is the standard deviation of the residuals. Residuals represent the distance between the regression line and the data points.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}},$$

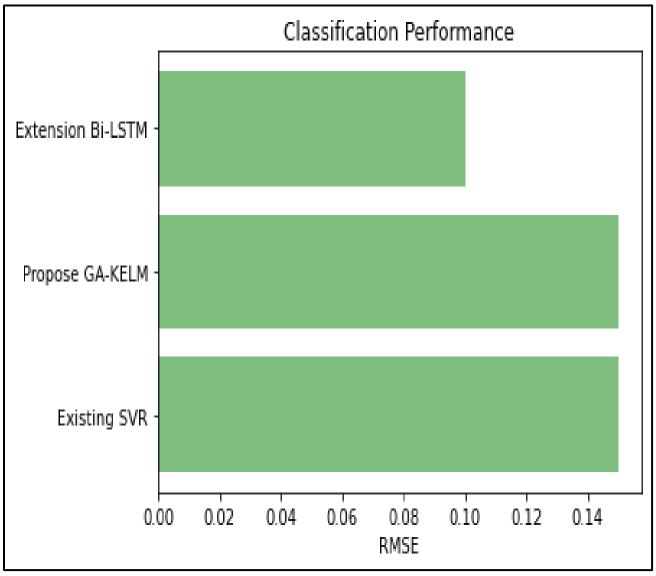

Fig 3 RMSE Comparison Graph

➢ *MAE:*

Absolute Error is the amount of error in your measurements. It is the difference between the measured value and "true" value. For example, if a scale states 90 pounds but you know your true weight is 89 pounds, then the scale has an absolute error of 90 lbs – 89 lbs = 1 lbs.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i |$$

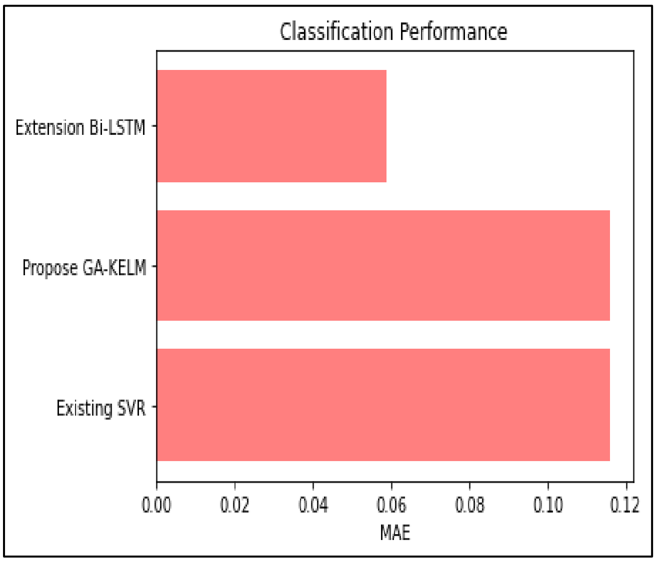

Fig 4 MAE Comparison Graph

Table 2 Performance Evaluation Table

	ML Model	MSE	RMSE	MAE
0	Existing SVR	0.023	0.15	0.116
1	Propose GA-KELM	0.023	0.15	0.116
2	**Extension Bi-LSTM**	**0.010**	**0.10**	**0.059**

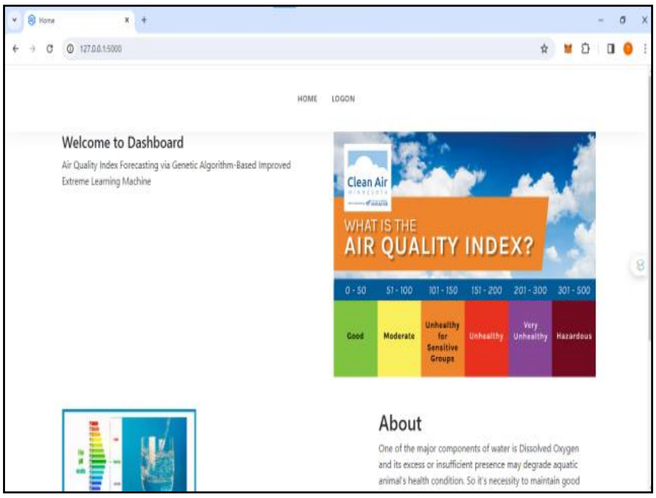

Fig 5 Home Page

Fig 6 Registration Page

Fig 7 Login Page

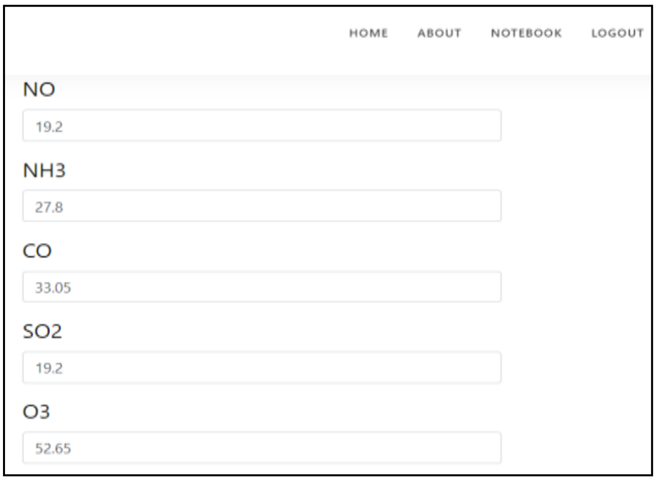

Fig 8 Upload Input Data

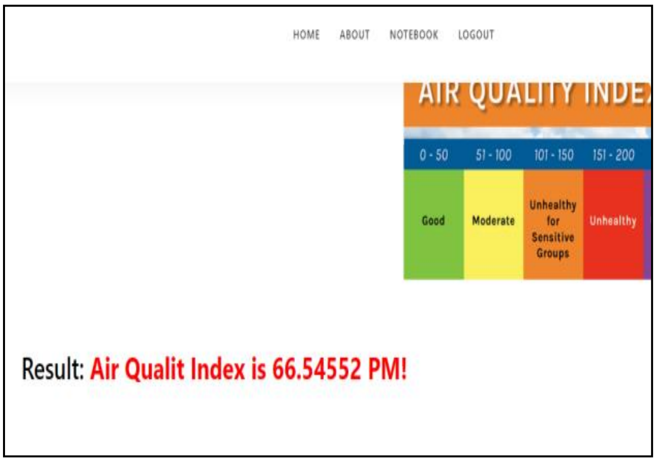

Fig 9 Final Outcome

V. CONCLUSION

In conclusion, the integration of Genetic Algorithm with Extreme Learning Machine (GA-KELM)[14] and the extension with Bidirectional Long Short-Term Memory (BiLSTM) represent significant advancements in air quality prediction, offering improved accuracy and enhancing environmental management decision-making. The deployment of the BiLSTM model within a user-friendly Flask framework further extends the project's impact, providing practical access to air quality predictions for researchers and the public alike. This not only empowers individuals to make informed decisions for their health and well-being but also facilitates proactive measures to mitigate the adverse effects of air pollution on the environment.

FUTURE SCOPE

Looking to the future, there are several avenues for further research and development. Firstly, continued refinement and optimization of the GA-KELM and BiLSTM models could lead to even greater predictive accuracy and robustness. Additionally, integrating real-time data streams and incorporating more diverse features into the models may enhance their capabilities further. Furthermore, exploring the application of these models in other domains beyond air quality prediction, such as climate modeling or environmental impact assessments, could yield valuable insights. Finally, efforts to enhance accessibility and usability of air quality prediction tools, including mobile applications and web-based platforms, can facilitate broader adoption and impact in addressing environmental challenges. Overall, the future scope lies in advancing computational techniques and leveraging them effectively to address evolving environmental concerns and improve quality of life globally.

REFERENCES

[1]. X. Li, L. Jin, and H. Kan, ''Air pollution: A global problem needs local fixes,'' Nature, vol. 570, no. 7762, pp. 437–439, Jun. 2019.

[2]. Y. Han, J. C. K. Lam, and V. O. K. Li, ''A Bayesian LSTM model to evaluate the effects of air pollution control regulations in China,'' in Proc. IEEE Big Data Workshop (Big Data), Dec. 2018, pp. 4465–4468.

[3]. L. Bai, J. Wang, X. Ma, and H. Lu, ''Air pollution forecasts: An overview,'' Int. J. Environ. Res. Public Health, vol. 15, no. 4, p. 780, 2018.

[4]. Y. Ding and Y. Xue, ''A deep learning approach to writer identification using inertial sensor data of air-handwriting,'' IEICE Trans. Inf. Syst., vol. E102-D, no. 10, pp. 2059–2063, 2019.

[5]. S.-Q. Dotse, M. I. Petra, L. Dagar, and L. C. De Silva, ''Application of computational intelligence techniques to forecast daily PM10 exceedances in Brunei Darussalam,'' Atmos. Pollut. Res., vol. 9, no. 2, pp. 358–368, Mar. 2018.

[6]. M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, ''Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications,'' Automat. Construct., vol. 101, pp. 111–126, May 2019.

[7]. S. Abirami, P. Chitra, R. Madhumitha, and S. R. Kesavan, ''Hybrid spatio-temporal deep learning framework for particulate matter (PM2.5) concentration forecasting,'' in Proc. Int. Conf. Innov. Trends Inf. Technol. (ICITIIT), Feb. 2020, pp. 1–6.

[8]. Y. Cheng, S. Zhang, C. Huan, M. O. Oladokun, and Z. Lin, ''Optimization on fresh outdoor air ratio of air conditioning system with stratum ventilation for both targeted indoor air quality and maximal energy saving,'' Building Environ., vol. 147, pp. 11–22, Jan. 2019.

[9]. A. C. Cosma and R. Simha, ''Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions,'' Building Environ., vol. 148, pp. 372–383, Jan. 2019.

[10]. M. Bhowmik, K. Deb, A. Debnath, and B. Saha, ''Mixed phase Fe2O3/Mn3O4 magnetic nanocomposite for enhanced adsorption of methyl orange dye: Neural network modeling and response surface methodology optimization,'' Appl. Organometallic Chem., vol. 32, no. 3, p. e4186, Mar. 2018.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2035

[11]. V. Chaudhary, A. Deshbhratar, V. Kumar, and D. Paul, ‘‘Time series based LSTM model to predict air pollutant's concentration for prominent cities in India,'' in Proc. Int. Workshop Utility-Driven Mining (UDM), Aug. 2018, pp. 1–9.

[12]. M. Chen, J. Yang, L. Hu, M. S. Hossain, and G. Muhammad, ‘‘Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators,'' IEEE Commun. Mag., vol. 56, no. 11, pp. 14–20, Nov. 2018.

[13]. R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, ‘‘A hybrid CNN-LSTM model for typhoon formation forecasting,'' GeoInformatica, vol. 23, no. 3, pp. 375–396, Jul. 2019.

[14]. S. Du, T. Li, Y. Yang, and S. Horng, ‘‘Deep air quality forecasting using hybrid deep learning framework,'' IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2412–2424, Jun. 2021.

[15]. R. Feng, H.-J. Zheng, H. Gao, A.-R. Zhang, C. Huang, J.-X. Zhang, K. Luo, and J.-R. Fan, ‘‘Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China,'' J. Cleaner Prod., vol. 231, pp. 1005–1015, Sep. 2019.

[16]. B. S. Freeman, G. Taylor, B. Gharabaghi, and J. Thé, ‘‘Forecasting air quality time series using deep learning,'' J. Air Waste Manage. Assoc., vol. 68, no. 8, pp. 866–886, Aug. 2018.

[17]. S. Mahajan, H.-M. Liu, T.-C. Tsai, and L.-J. Chen, ‘‘Improving the accuracy and efficiency of PM2.5 forecast service using cluster-based hybrid neural network model,'' IEEE Access, vol. 6, pp. 19193–19204, 2018.

[18]. J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, ‘‘An information framework for creating a smart city through Internet of Things,'' IEEE Internet Things J., vol. 1, no. 2, pp. 112–121, Apr. 2014.

[19]. A. Grover, A. Kapoor, and E. Horvitz, ‘‘A deep hybrid model for weather forecasting,'' in Proc. 21st ACM SIGKDD Int. Conf. Knowl. DiscoveryData Mining, Aug. 2015, p. 379–386.

[20]. A. Agarwal and M. Sahu, ‘‘Forecasting PM2.5 concentrations using statistical modeling for Bengaluru and Delhi regions,'' Environ. Monit. Assessment, vol. 195, p. 502, Mar. 2023.

[21]. D. J. Lary, T. Lary, and B. Sattler, ‘‘Using machine learning to estimate global PM2.5 for environmental health studies,'' Environ. Health Insights, vol. 9, no. 1, pp. 41–52, 2015.

[22]. Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, ‘‘Forecasting fine-grained air quality based on big data,'' in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, Aug. 2015, pp. 2267–2276.

[23]. Y. Zheng, L. Capra, O. Wolfson, and H. Yang, ‘‘Urban computing: Concepts, methodologies, and applications,'' ACM Trans. Intell. Syst. Technol., vol. 5, no. 3, p. 38, Sep. 2014.

[24]. T. S. Rajput and N. Sharma, ‘‘Multivariate regression analysis of air quality index for Hyderabad city: Forecasting model with hourly frequency,'' Int. J. Appl. Res., vol. 3, no. 8, pp. 443–447, 2017.

[25]. Z. Kang and Z. Qu, ‘‘Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou,'' in Proc. IEEE Comput. Intell. Appl. (ICCIA), Sep. 2017, pp. 155–160, doi: 10.1109/CIAPP.2017.8167199.

[26]. B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, ‘‘A sequence-tosequence air quality predictor based on the n-step recurrent prediction,'' IEEE Access, vol. 7, pp. 43331–43345, 2019.

[27]. K. Elbaz, I. Hoteit, W. M. Shaban, and S.-L. Shen, ‘‘Spatiotemporal air quality forecasting and health risk assessment over smart city of NEOM,'' Chemosphere, vol. 313, Feb. 2022, Art. no. 137636.

[28]. P. C. Campbell, Y. Tang, P. Lee, B. Baker, D. Tong, R. Saylor, A. Stein, J. Huang, H.-C. Huang, E. Strobach, J. McQueen, L. Pan, I. Stajner, J. Sims, J. Tirado-Delgado, Y. Jung, F. Yang, T. L. Spero, and R. C. Gilliam, ‘‘Development and evaluation of an advanced national air quality forecasting capability using the NOAA Global Forecast System version 16,'' Geosci. Model Develop., vol. 15, no. 8, pp. 3281–3313, Apr. 2022.

[29]. X.-B. Jin, Z.-Y. Wang, W.-T. Gong, J.-L. Kong, Y.-T. Bai, T.-L. Su, H.-J. Ma, and P. Chakrabarti, ‘‘Variational Bayesian network with information interpretability filtering for air quality forecasting,'' Mathematics, vol. 11, no. 4, p. 837, Feb. 2023.

[30]. J. Li, W. Shi, and D. Yang, ‘‘Color difference classification of dyed fabrics via a kernel extreme learning machine based on an improved grasshopper optimization algorithm,'' Color Res. Appl., vol. 46, no. 2, pp. 388–401, Apr. 2021.